On the Suppression of Noise from a Fast Moving Acoustic Source using Multimodality

Wendyam Serge Boris Ouedraogo, Bertrand Rivet, Christian Jutten

On the Suppression of Noise from a Fast Moving Acoustic Source using Multimodality

Wendyam Serge Boris Ouedraogo*, Bertrand Rivet, and Christian Jutten

Université Grenoble-Alpes, GIPSA-lab, F-38000 Grenoble, France
{wendyam.ouedraogo,bertrand.rivet,christian.jutten}@gipsa-lab.grenoble-inp.fr

Abstract. The problem of cancelling the noise from a moving acoustic source in outdoor environment is investigated in this paper. By making use of the known instantaneous location of the moving source (provided by a second modality), we propose a time-domain method for removing the noise from a moving source in a mixture of acoustic sources. The proposed method consists in resampling the mixed data recorded at a reference sensor, and by linearly combining the resampled data and the non-resampled data of the others sensor to cancel the undesired source. Simulation on synthetic data show the effectiveness and the usefulness of the proposed method.

Keywords: Non-stationary noise suppression, Moving acoustic source, Multimodality, Delay compensation, Beamforming

1 Introduction

The task of removing noise of a dominant undesired source in a mixture of several acoustic sources has application in many areas, like hearing aids, target acoustic source enhancement in wireless communication, or auditory scene analysis to name a few [1][2]. This issue remains a challenge when the undesired source is moving. In this paper, we consider the problem where a target fixed acoustic source is corrupted by a dominant noise from a fast moving acoustic source. We consider the case where the moving source moves along a straight line and at a constant speed in a plane and we are interested by removing this undesired source. The high speed of the moving source may cause a doppler frequency shift between the recording microphones [3]. The problem is discussed in an outdoor environment and we do not take into account the reflexion over the ground. This kind of problem can occur in airport platforms, on roadsides or along railway lines. The problem described above belongs to general framework of non-stationary noise suppression. Conventional methods for suppression of non stationary noise include spectral subtraction, source separation and beamforming [4][5][6].

The spectral subtraction based methods performs subtraction of the estimated noise spectrum to the spectrum of the recorded signal (containing both

* This work was supported by the project CHESS, 2012-ERC-AvG-320684

the target source and the noise) to get an estimated noiseless spectrum of the target acoustic source from which one can compute the temporal profile of the target source [4][7]. However, these methods require good approximations of the noise spectrum, which are difficult to get in the case of highly non-stationary noise, as noise due to a fast moving source.

Another approach to remove undesired source from the mixture is to perform source separation. Most of source separation methods have been proposed for fixed sources and exploit the source statistics, especially their mutual independence. For moving sources, some methods assume that the sources are moving slowly enough so that one can divide the mixed data into short periods in which the sources can be assumed fixed [5][8], the separation is then performed blockwise, using statistical methods. However, this assumption no longer holds if the sources are moving faster so that the period for which they can be considered as fixed is not enough to perform statistical processing.

Regarding the beamforming methods, they proceed by combination of the signals recorded on the different sensors in order to achieve a spatial filtering that will pass the signal from a direction of interest (direction of the target source) and remove noises from all other directions [9]. Recently, new methods coupling video and audio modalities have been proposed to unmix moving acoustic sources [10][11][12]. These techniques firstly estimate the instantaneous positions and the velocities of the sources by the video, and make use of the source locations to design beamforming method to estimate the source temporal profiles.

The method proposed in this paper is within the scope of multimodality methods, and we assume that a second modality (other than the audio) like video or Global Positioning System (GPS) provides the location of the moving source at each instant. The method consists in a first step to resample the recorded data of a reference sensor in order to compensate the difference of propagation delays of the undesired source between a given sensor and the reference sensor. In a second step, we linearly combine the resampled data and the non-resampled data of the other sensors to remove the undesired source. The paper is organized as follows: Section 2 presents the problem modelization and a simple way to instantaneously estimate the location of the moving source. Section 3 presents the proposed method for removing an undesired and dominant fast moving source. In Section 4, we show simulation results and in Section 5 we derive conclusions.

2 Mixing model and estimation of source locations

The mixing model is given by equation (1):

$$m_k(t) = a_{k1} s_1\left(t - \tau_{k1}\right) + a_{k2}\left(t - \tau_{k2}(t)\right) s_2\left(t - \tau_{k2}(t)\right), \; 1 \leq k \leq K. \quad (1)$$

where $m_k(t)$ is the recorded mixed data at sensor k, K is the total number of sensors. $s_1(t)$ and $s_2(t)$ are respectively the target fixed source and the undesired moving one. $a_{kl}(t)$ and $\tau_{kl}(t)$ are the mixing coefficient and the propogation delay from source l to sensor k at time index t. Since $s_1(t)$ is fix, then $a_{k1}(t) = a_{k1}$

and $\tau_{k1}(t) = \tau_{k1}$. The delay $\tau_{k2}(t)$ between source 2 and sensor k at time index t is proportional to the distance $D_{k2}(t)$ between source 2 and sensor k and in a far field context, we can assume that the mixing coefficent $a_{k2}(t)$ is inversely proportional to the distance $D_{k2}(t)$. Denoting c the sound velocity in air, it leads to:

$$a_{k2}(t) = \frac{1}{D_{k2}(t)} \text{ and } \tau_{k2}(t) = \frac{D_{k2}(t)}{c}. \tag{2}$$

Let $(X_{m_k}, Y_{m_k}, Z_{m_k})$ be the known location of the sensor m_k, and the instantaneous location of the moving source s_2 be $(X_{s_2}(t), Y_{s_2}(t), Z_{s_2}(t))$. The distance between source 2 and sensor k at time t is given by:

$$D_{k2}(t) = \sqrt{(X_{m_k} - X_{s_2}(t))^2 + (Y_{m_k} - Y_{s_2}(t))^2 + (Z_{m_k} - Z_{s_2}(t))^2}. \tag{3}$$

For a reference time index t_0, the location of the source s_2 is given by the coordinates $(X_{s_2}(t_0), Y_{s_2}(t_0), Z_{s_2}(t_0))$. Since the source s_2 is moving along a straight line at a constant speed v_2 in a plane, for example in the XY plane (see Fig. 1), the coordinates of s_2 at any time t is given by:

$$\begin{cases} X_{s_2}(t) = X_{s_2}(t_0) + v_2(t - t_0)\cos(\theta_2) \\ Y_{s_2}(t) = Y_{s_2}(t_0) + v_2(t - t_0)\sin(\theta_2) \\ Z_{s_2}(t) = Z_{s_2}(t_0) \end{cases} \tag{4}$$

where θ_2 is the angle between the x-axis and the axis of the moving source (Fig. 1). The distance $D_{k2}(t)$ is then given by:

$$D_{k2}(t) = \sqrt{D_{k2}^2(t_0) - 2v_l(t - t_0)\left[(X_{m_k} - X_{s_2}(t_0))\cos(\theta_2) + (Y_{m_k} - Y_{s_2}(t_0))\sin(\theta_2)\right] + v_2^2\,(t - t_0)^2}. \tag{5}$$

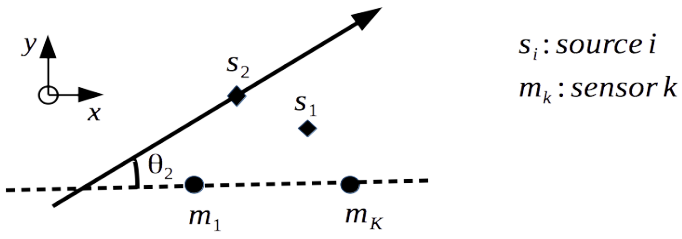


Fig. 1. Mixing configuration

Equations (4) and (5) show that estimation of the current localization of s_2, and of the distance $D_{k2}(t)$ between source 2 and sensor k, only requires the position of s_2 at a reference time t_0, the speed v_2 of s_2 and the angle θ_2. These parameters can be obtained with a second modality like GPS in a cooperative environment, or estimated by processing video and/or audio signal [13][14]. In

this paper, we will not address the problem of getting the previous parameters of the moving source, but we assume that they are known from the methods mentioned above. Furthermore, since θ_2 is constant and without loss of generality, we will consider in the remaining of the paper that $\theta_2 = 0$, to simplify the equations. In that case, $D_{k2}(t)$ is given by:

$$D_{k2}(t) = \sqrt{D_{k2}^2(t_0) - 2v_2(t-t_0)\left[(X_{m_k} - X_{s_2(t_0)})\right] + v_2^2\,(t-t_0)^2}. \tag{6}$$

3 Suppression of the undesired moving source

Below we describe a method for removing the fast source s_2, in this purpose, we choose a reference sensor r. The mixed data recorded at the sensor r and at any sensor k are respectively given by:

$$m_r(t) = a_{r1}s_1\,(t-\tau_{r1}) + a_{r2}\,(t-\tau_{r2}(t))\,s_2\,(t-\tau_{r2}(t)) \tag{7}$$

$$m_k(t) = a_{k1}s_1\,(t-\tau_{k1}) + a_{k2}\,(t-\tau_{k2}(t))\,s_2\,(t-\tau_{k2}(t)) \tag{8}$$

According to equations (6) and (2), $\tau_{r2}(t)$ can be written as:

$$\tau_{r2}(t) = \sqrt{\tau_{r2}^2(t_0) \; - \frac{2v_2(t-t_0)\left(X_{m_r} - X_{s_2(t_0)}\right)}{c^2} + \frac{v_2^2\,(t-t_0)^2}{c^2}}. \tag{9}$$

The first step of the proposed method consist to irregularly resample the signal received on sensor r, to compensate the time difference of arrival of source s_2 between sensor r and sensor k. The resampled mixed data is given by:

$$\begin{aligned} m_r(t+\varepsilon_{kr}(t)) = {} & a_{r1}s_1\left[t+\varepsilon_{kr}(t)-\tau_{r1}\right] \\ & + a_{r2}\left[t+\varepsilon_{kr}(t)-\tau_{r2}[t+\varepsilon_{kr}(t)]\right] s_2\left[t+\varepsilon_{kr}(t)-\tau_{r2}[t+\varepsilon_{kr}(t)]\right] \end{aligned} \tag{10}$$

where $\varepsilon_{kr}(t)$ is a time shift that allows equalizing the delays of arrival of source 2 on sensors r and k at time t. We then seek $\varepsilon_{kr}(t)$ such as:

$$\varepsilon_{kr}(t) - \tau_{r2}[t+\varepsilon_{kr}(t)] = -\tau_{k2}(t). \tag{11}$$

One can easily show that:

$$\tau_{r2}[t+\varepsilon_{kr}(t)] = \sqrt{\tau_{r2}^2(t) \; + \frac{v_2^2}{c^2}[\varepsilon_{kr}(t)]^2 + \frac{2v_2\varepsilon_{kr}(t)}{c^2}\left[v_2(t-t_0) - (X_{m_r} - X_{s_2(t_0)})\right]}. \tag{12}$$

Otherwise:

$$\varepsilon_{kr}(t) - \tau_{r2}[t+\varepsilon_{kr}(t)] = -\tau_{k2}(t) \Rightarrow \left[\tau_{k2}(t) + \varepsilon_{kr}(t)\right]^2 = \tau_{r2}[t+\varepsilon_{kr}(t)]^2. \tag{13}$$

By combining equation (12) and equation (13), one gets the quadratic equation (14), for which one of the two solutions gives the desired $\varepsilon_{kr}(t)$.

$$[\varepsilon_{kr}(t)]^2\left[1 - \frac{v_2^2}{c^2}\right] + 2\varepsilon_{kr}(t)\left[\tau_{k2}(t) - \frac{v_2^2(t-t_0) - v_2(X_{m_r} - X_{s_2(t_0)})}{c^2}\right] + \left[\tau_{k2}^2(t) - \tau_{r2}^2(t)\right] = 0. \tag{14}$$

It follows that:

$$m_r(t+\varepsilon_{kr}(t)) = a_{r1}s_1\left[t-\tau_{r1}+\varepsilon_{kr}(t)\right] + a_{r2}\left[t-\tau_{k2}(t)\right]s_2\left[t-\tau_{k2}(t)\right]. \quad (15)$$

Thus, for removing the moving source s_2 from sensor k, we just have to linearly combine $m_k(t)$ and $m_r(t+\varepsilon_{kr}(t))$ as illustrated on equation (16):

$$\begin{aligned}\tilde{m}_k(t) &= a_{r2}\left[t-\tau_{k2}(t)\right]m_k(t) - a_{k2}\left[t-\tau_{k2}(t)\right]m_r(t+\varepsilon_{kr}(t)) \\ &= a_{k1}a_{r2}\left[t-\tau_{k2}(t)\right]s_1(t-\tau_{k1}) - a_{r1}a_{k2}\left[t-\tau_{k2}(t)\right]s_1\left[t-\tau_{r1}+\varepsilon_{kr}(t)\right].\end{aligned} \quad (16)$$

By construction, $\tilde{m}_k$ is function of s_1 only, thus a scaled and delayed estimation $\hat{s}_1$ of the target source s_1 can be estimated by:

$$\hat{s}_1(t) = a_{k1}s_1(t-\tau_{k1}) - a_{r1}\frac{a_{k2}\left[t-\tau_{k2}(t)\right]}{a_{r2}\left[t-\tau_{k2}(t)\right]}s_1\left[t-\tau_{r1}+\varepsilon_{kr}(t)\right]. \quad (17)$$

The method described above allows to remove the undesired moving source, but it also creates a second path of the target source (rigth term in (17)), that can be dealed with multipaths suppression algorithms.

4 Simulation Results

This section presents simulation results on synthetic data. The proposed method is compared to the Linearly Constrained Minimum Variance (LCMV) beamforming method [15]. Let's recall that LCMV-beamforming performs a spatial filtering that pass the signal from a given direction of interest while minimizing noise from all other directions. The efficiency of the estimation of s_1 is quantified by the signal-to-interference ratio (SIR) and by the signal-to-distortion ratio (SDR), as defined in [16]. To compute the SIR and SDR, $\hat{s}_1$ is decomposed as:

$$\hat{s}_1 = s_{target} + e_{interf} + e_{artifact}. \quad (18)$$

where s_{target} is a scaled and delayed version of the original source s_1, and where e_{interf}, and $e_{artifact}$ are the interference, and artifact error terms, respectively [16]. The SIR and the SDR are then computed through equation (19) and equation (20), respectively. The larger the SIR and the SDR are, the better the estimation is.

$$SIR = 10\log_{10}\frac{\|s_{target}\|^2}{\|e_{interf}\|^2}. \quad (19)$$

$$SDR = 10\log_{10}\frac{\|s_{target}\|^2}{\|e_{interf}+e_{noise}+e_{artifact}\|^2}. \quad (20)$$

The fixed source, s_1, is a speech while the moving source, s_2, is a tone at frequency $300Hz$ that moves at $50km/h$ parallel to the x-axis. Fig. 3 shows the original source s_1 and the mixed signal recorded at microphone 1. We set the coordinates of s_1 to $(0m, 0.1m, 1.5m)$ and the coordinates of s_2 at the reference

time t_0 to $(0m, 5m, 0m)$. We consider a uniform linear antenna whose sensors are distributed along the x-axis and centered at the origin of this axis. The distance between two consecutive sensors is $d = 5cm$ and for each sensor k, we set $Y_{m_k} = 0$ and $Z_{m_k} = 1.5m$. Fig. 2 shows the simulation scenario.

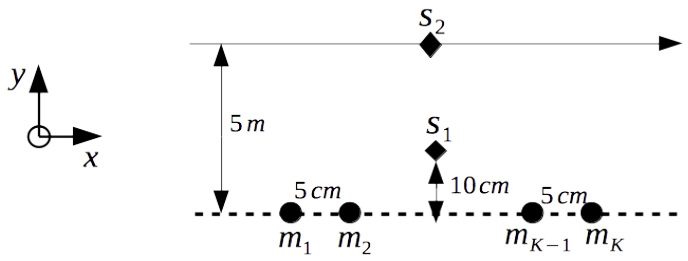
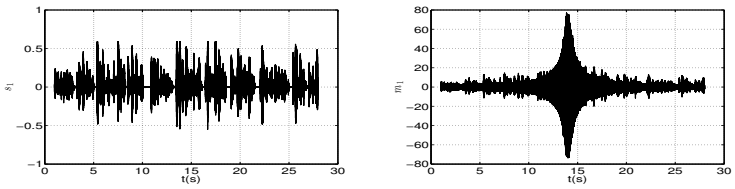


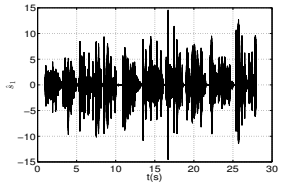
Fig. 2. Simulation scenario



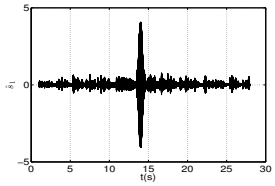
(a) Original source s_1 (b) Mixed data at microphone 1

Fig. 3. Original source s_1 and mixed data at microphone 1

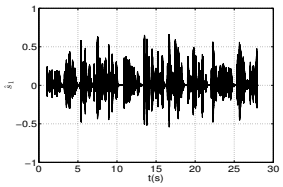
Fig. 4 shows the estimation of the target source, $\hat{s}_1$ after the removing of the undesired moving source, by proposed method and by beamforming. Table 1 shows the performance indices of the proposed method and of LCMV-Beamforming versus K, the number of microphones.
One can see from Fig. 4 that the proposed method is able to completely remove the undesired moving source with only 2 microphones and there is no need to increase the number of sensors. To achieve this performance the LCMV-beamforming method requires at least 3 microphones (in this example). Regarding the performance reported in Table 1, our method is better than the LCMV-beamforming for 2 microphones, and the SIR is as good as those of LCMV-beamforming with a large number of sensors. However, the signal to distortion ratio estimated by the proposed method is quite low and must be improved. This could be done by developping a post-processing method (which could be inspired by echo canceller methods) that will remove the second path generated by our method.



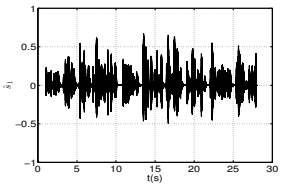
(a) Proposed method with 2 microphones



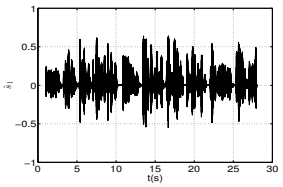
(b) LCMV-Beamforming with 2 microphones



(c) LCMV-Beamforming with 3 microphones



(d) LCMV-Beamforming with 4 microphones



(e) LCMV-Beamforming with 5 microphones

Fig. 4. Results of estimation of the target source s_1

	Proposed method	LCMV-Beamforming			
	$K = 2$	$K = 2$	$K = 3$	$K = 4$	$K = 5$
$SIR(dB)$	61.74	11.88	63.83	60.42	64.16
$SDR(dB)$	-0.64	-14.31	20.22	19.24	33.67

Table 1. Performance indices

5 Conclusion

In this paper, we propose a method for removing a dominant and fast moving undesired acoustic source. The proposed method requires only two microphones and exploits the known position of the moving source, assumed to be provided by other modality like video or GPS. It consists firstly in resampling the recorded data of a reference sensor in order to compensate the delay difference of the undesired source between a given sensor and the reference sensor. After that we linearly combine the resampled data and the non-resampled data of the second sensor for removing the undesired source. Simulation on synthetic data shows that the proposed method outperforms beamforming with 2 sensors. Future works include developping a post-processing method to remove the second path of the target source created by our method in order to improve the output signal to distorsion ratio. The case of multiple undesired moving sources will also be investigated in a near future.

References

1. Widrow, B. , Luo, F.-L.: Microphone Arrays for Hearing Aids: An overview. Speech Communication, Vol. 39, 139-146 (2003)
2. Okutani, K., Yoshida, T., Nakamura, K., Nakadai, K.: Outdoor Auditory Scene Analysis Using a Moving Microphone Array Embedded in a Quadrocopter. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, 3288–3293 (2012)
3. Australian Centre for Field Robotics, Chapter 14 - Doppler Measurement. `http://www.acfr.usyd.edu.au/pdfs/training/sensorSystems/14%20Doppler%20Measurement.pdf`
4. Manohar, K., Rao, P.: Speech enhancement in nonstationary noise environments using noise properties. Speech Communication, Vol. 48, 96-109 (2006)
5. Anemuller, J., Gramss, T.: On-line Blind Separation of Moving Sound Sources. In: ICA'99 (1999)
6. Farrell, K., Mammone, R., Flanagan, J.L.: Beamforming Microphone Arrays for Speech Enhancement. In: IEEE International Conference on Acoustics, Speech, and Signal Processing, 285–288 (1992)
7. Boll, S.: Suppression of acoustic noise in speech using spectral subtraction. IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. 27, NO. 2, 113–120 (1979)
8. Mukai, R., Sawada, H., Haraki, S., Makino, S.: Blind Source Separation for Moving Speech Signals Using Blockwise ICA and Residual Crosstalk Subtraction. IEICE Trans. Fundamentals, VOL.E87A, NO.8, 1941–1948 (2004)
9. Van Veen, B., Buckley, K.: Beamforming: A Versatile Approach to Spatial Filtering. IEEE ASSP Mag.,Vol 5, NO. 2, 4–24 (1998)
10. Maganti, H.K., Gatica-Perez, D., McCowan, I.: Speech Enhancement and Recognition in Meetings With an Audio-Visual Sensor Array. IEEE Trans. on Audio, Speech, and Language Processing, Vol. 15, NO. 8, 2257–2269 (2007)
11. Naqvi, S.M., Yu, M., Chambers, J.A.: A Multimodal Approach to Blind Source Separation of Moving Sources. IEEE Journal of Selected Topics in Signal Processing, Vol. 4, NO. 5, 895–910 (2010)
12. Naqvi, S.M., Wang, W., Khan, M.S., Barnard, M., Chambers, J.A.: Multimodal (Audiovisual) Source Separation Exploiting Multi-speaker Tracking, Robust Beamforming and Time-Frequency Masking. Signal Processing, IET, Vol. 6, NO. 5, 466–477 (2012)
13. Strobel, N., Spors, S., Rabenstein, R.: Joint audio-video object localization and tracking. IEEE Signal Processing Magazine, Vol. 18, NO. 1, 22–31 (2001)
14. Nishie, S., Akagi, M.: Acoustic sound source tracking for a moving object using precise Doppler-Shift measurement. In Proc. of the 21st European Signal Processing Conference, 1–5 (2013)
15. Souden, M., Benesty, J., Affes, S.: A Study of the LCMV and MVDR Noise Reduction Filters. IEEE Transactions on Signal Processing, Vol 58, NO. 9, 4925–4935 (2010)
16. Vincent, E., Gribonval, R., Fevotte, C.: Performance measurement in blind audio source separation. IEEE Transactions on Audio, Speech, and Language Processing, Vol. 14, NO. 4, 1462-1469 (2006)